ONO5425

92-06926/

Return to
OISE

1050-85

A1

The Differentiation of Knowledge and Learning in an Educational Achievement Survey[1]

Richard G. Wolfe
The Ontario Institute for Studies in Education
Toronto, Canada

The Second IEA Mathematics Study (SIMS) included a 'Longitudinal, Classroom-Process' component, which provides a first instance of an international achievement survey that has been carried out with student measurement at more than one point of time. In fact, there are few examples of national or subnational studies that have this important feature. The crucial advantage of a longitudinal design is that it allows differentiating:

- the *knowledge* that students have acquired by some point in the school system, and the host of factors that lead to the knowledge, from
- the *learning* that takes place in a specific period of school, and the educational factors that determine the learning.

The purposes of this paper are to outline the design and execution of the longitudinal component of SIMS, to describe and illustrate the kinds of analysis and policy inference that are being made with the data, and to suggest the importance of the longitudinal component for future studies and propose some extensions of the methodology.

The SIMS Longitudinal Design

The Second IEA Mathematics Study was carried out in twenty countries and at two levels of schooling: Population A corresponds to the school grade in which the modal student age is 13 years, and Population B corresponds to students specializing in mathematics in the last year of secondary education. The longitudinal component was applied only to Population A, which in the

---

1. Presentation to the General Assembly of the International Association for Evaluation of Educational Achievement, New York, September 13, 1987

U.S.A. and Canada means grade 8, and only to eight of the countries[2]: Flemish Belgium, British Columbia, Ontario, France, Japan, New Zealand, Thailand, and U.S.A.

For these countries, students were given the mathematics tests twice: at the beginning of the school year and again at the end. This is the *longitudinal* dimension. The same students, classrooms, and schools participated at the beginning-of-year and end-of-year testing. The survey was based on a sample of *classrooms*. The total array of instrumentation and data collection was as follows:

- At the beginning of the school year, students filled out a very brief background and attitude questionnaire and took a version of the mathematics test. This is the *pretest*.

- During the school year, the teachers of the sampled classrooms filled out a set of questionnaires that detailed their teaching of the different mathematics topics as those arose.

- At the end of the school year, the students answered a long background and attitude questionnaire and took the same or another version of the mathematics test. This is the *posttest*.

- Also, at the end of the year, the teachers filled out a background questionnaire, a questionnaire about their general teaching practices, and a complete mapping of their estimates of their students' opportunities to learn the contents corresponding to the mathematics test items. Also, the school principal and the mathematics department head filled out a general information questionnaire about the school and particularly the organization of the mathematics program.

The full study design called for a 40-item *core* test that all students took at the beginning and the end of the year, plus four 35-item *rotated* forms, and each student took one of those forms, randomly assigned within classroom, at the beginning of the year and the same or another of the forms at the end of the year, with randomized rotation of all pairs of forms. In total, 180 test items appeared in the core and rotated forms, and all items were used at both pretest and posttest. The exact nature of the implementation varied somewhat from country to country.[3]

---

2. The term 'countries' means, as can be seen, a whole nation or a separate but complete educational jurisdiction within a nation. A similar survey has since been carried out in the Dominican Republic. In the U.S.A., a longitudinal component was included additionally for its SIMS Population B survey.

3. In Japan, which did the study a year earlier, there were 176 items, divided into a 40-item core and four 34-item rotated forms. Also, the Japanese pretest consisted of a special 60-item selection from the 176. In the other seven

During the organizational phases of SIMS, there was considerable doubt expressed about the feasibility of making a longitudinal data collection, but except in one country where identification code linkage was lost, there were no special data processing problems associated with the longitudinal nature of the study. The costs, of course, were higher, but in terms of a total budget not greatly higher, since data collection is a relatively small expense, and data processing costs vary mainly with programming and analysis and not data volumes.

Analytic Perspectives and Applications

The primary purpose of an educational achievement survey is to *describe the distribution of achievement*, and the descriptive analysis is aimed at:

- *content*, in terms of specific and global areas of the subject area and level of cognitive complexity, and
- *population*, differentiating by student background, school categories, and geographical and national divisions.

In a cross-sectional study, the dependent variable of educational achievement represents a point in the cumulative acquisition of *knowledge*. If we are testing grade 8 students, there is no sense of when an achievement was obtained: it could as well be grade 5, 6, or 7 as grade 8. For this reason, perhaps, educational assessments surveys are often made at critical junctures in school systems, such as final or transitional years. The SIMS Population A was chosen because for most countries it represents the last year when all the age cohort is in school and taking a common mathematics curriculum.[4] The goal may be to evaluate and compare the *products* of the school systems.

The perspective added by the longitudinal measurement component is information, detailed across contents and populations, about the *learning* that takes place during, in our case, one particular school year. While description of the accumulated knowledge distribution may be appropriate for evaluating

---

countries, there were 180 items, with 157 in common with the Japanese set. The division of these into the 40-item core and 35-item rotated forms was allowed to vary somewhat, to allow more nationally relevant content to appear in the core test. In British Columbia, only the core was used in the pretesting. There were different modes of connecting the randomization of pretest and posttest rotated form: In Thailand, each student took the same form pre and post. In France, each student took a different form. In the other countries, there was systematic rotation to ensure that each combination of pre and post form was evenly distributed.

4. This is not really true of Flemish Belgium, where by this age there is already differentiation with different academic programs in different school types, nor in the U.S.A., where in grade 8 there is very strong within-school tracking in mathematics classes.

the products of an educational system, information about the distribution of *learning* is central to assessing its curriculum and teaching.

Display 1 gives a basic longitudinal description of achievement across countries and content areas and illustrates the added value of the longitudinal component.

- The dark areas below the reference line index the *pretest* achievement. We can see that for most countries, performance is relatively high at the beginning of the school year in the Fractions topic and low in Geometry.

- The black areas just above the reference line index *growth* – that is, pretest to posttest change. Although these test items were specifically written and selected to correspond to the mathematics curriculum of Population A, the overall result is that the growth rates are modest, with some exceptions: fractions in France, algebra in Japan. The dark plus the grey equals posttest achievement.

- The light areas at the top of each bar represents the non-achievement. In only a few cases, such as fractions for Japan, is there any lack of room for growth.

A parallel and explanatory picture appears as **Display 2**, which is based on the teachers' reports of student opportunity to learn the mathematics necessary to answer the test items.[5] We see great variation in the *implemented* curriculum. Very little geometry is considered prior content in any country. France, Japan, and Flemish Belgium stand out in having the much of the measurement and ratio-proportion-percent content appearing before the Population A grade. Japan is unusual in that nearly all content about fractions is regarded as have been taught previously.

By comparing the achievement and the opportunity to learn displays, we can conclude that reteaching old material where there is already a reasonable level of achievement yields little growth. For example, there is large growth in fractions only in France, where the pretest scores were low and the material was evidently new. This is, in fact, the point in the French curriculum for teaching fractions. In Ontario, which is typical of the other countries, fractions in grade 8 is mainly a review topic, and the (considerable) time spent on fractions yields little new achievement.

---

5. For each test item, the teachers were asked first what proportion of their students they thought could answer the questions correctly. Then they were asked whether they had taught or reviewed the 'mathematics necessary to answer the question' during the current school year. If they had not, they were asked why, and for this tabulation the key response is 'because it was taught in an earlier year.'

Relational Analysis

In any educational assessment survey such as SIMS, there is a certain enthusiasm for making relational analysis of the data, with a goal of explaining the student achievement results on the basis of student, teacher, and school factors. The theoretical and statistical difficulties of making such quasi-causal analysis from the non-experimental data base provided by an achievement survey are well known. Nevertheless, attempts will always be made, and in the case of SIMS, there have been especially high hopes for the longitudinal component because the mathematics pretest provides what has been generally lacking in previous studies: an adequate control variable for initial student input.

With the SIMS longitudinal data, we can demonstrate the dangers of relational analysis given *cross-sectional* data. Suppose that we measure opportunity to learn and achievement at the end of a school year. From the SIMS data for arithmetic in the U.S.A., the corresponding scatterplot and robust smoothed line is shown at the bottom of **Display 3**. Evidently, the more opportunity, the more achievement. But this relationship can be broken down temporally:

- For the beginning of the year, the pretest achievement of the students is compared with the estimates of prior opportunity to learn as shown at the top of the display. It can be seen that where initial achievement was higher, there was higher prior opportunity.

- For the end of the year, the achievement gain[6] is compared with the teachers' reports of opportunity to learn in the current year, as shown in the middle of the display. We see a trend, but not as pronounced.

Evidently, the end-of-year cross-sectional relationship represents the confounding of these processes; without the longitudinal data, we could not hope to disentangle them.

Design Issues and Recommendations

With the success of the longitudinal SIMS, both in its practicality and the evident value added to description and analysis, we should consider how to employ longitudinal methodology in future achievement surveys.

---

6. For this plot, gain is calculated as the residual in the regression of posttest on pretest, with classrooms as the unit of analysis.

More variables in the pretest. The SIMS pretest for six of the eight countries included all the 180 mathematics test items. This has proved enormously useful for purposes of both descriptive and relational analysis. For description, it has meant that growth can be pinpointed to very specific contents, which is important since the item by growth interactions are large, even within what was thought to be homogeneous subtest areas. For relational study, specific links can be built between pretest achievement in an area, opportunity to learn in that area, and resulting achievement. All items should be used at each time point.

Longitudinal attitude measurement. There is a small attitude scale in the SIMS pretest instrumentation, while the posttest questionnaire had a long and articulated attitude section. It has been very frustrating for the analysts not to have better description of the attitude change over the school year: are students expressing long-held attitudes toward mathematics, or are the attitudes (e.g., toward the creativity of mathematics study) being determined in this school year? We cannot tell. A full complement of attitude items should be included at each time point.

Number of time points. Since a two-time-point study has proved practical, it is natural consider longer-term studies. The substantive goal is to characterize the mathematics program of a whole level of schooling, such as the intermediate school or the high school. One would like a broader picture of content learning, to examine the effects of alternative orders of curricula (geometry or algebra first?) and the different amounts and placements of new material versus review or repetition. One would like to trace the growth of students as they pass through the school grades and get placed in tracks. An example of these kinds of results in given in **Display 4**, which shows achievement in a set of arithmetic items over four grades in Ontario.[7] There is good growth from grade 7 to 8. It is not clear from the diagram what happens in the split to tracks in grade 9. In grade 10, there is good achievement gain in each track.

Better measurement of opportunity to learn. The questions used for measuring opportunity to learn in SIMS are not very good: most importantly, they do not allow adequate differentiation of material that is *reviewed* from material that is *introduced*. A simple rewording would fix that. Another defect is that teachers may be poor judges of what content was introduced in previous

---

7. These are not SIMS data and in fact were obtained in a cross-sectional study. For many kinds of analysis and policy inference, a multi-grade sample, combined with a two-time-point data collection, might be very appropriate.

years,[8] but this kind of data would be obtained if the number of time points could be extended.

An Unresolved Problem

In conclusion, we should consider a research issue of great popular and political interest: the origin of gender differences in mathematics achievement. New insights into this field might require an amplified longitudinal achievement survey.

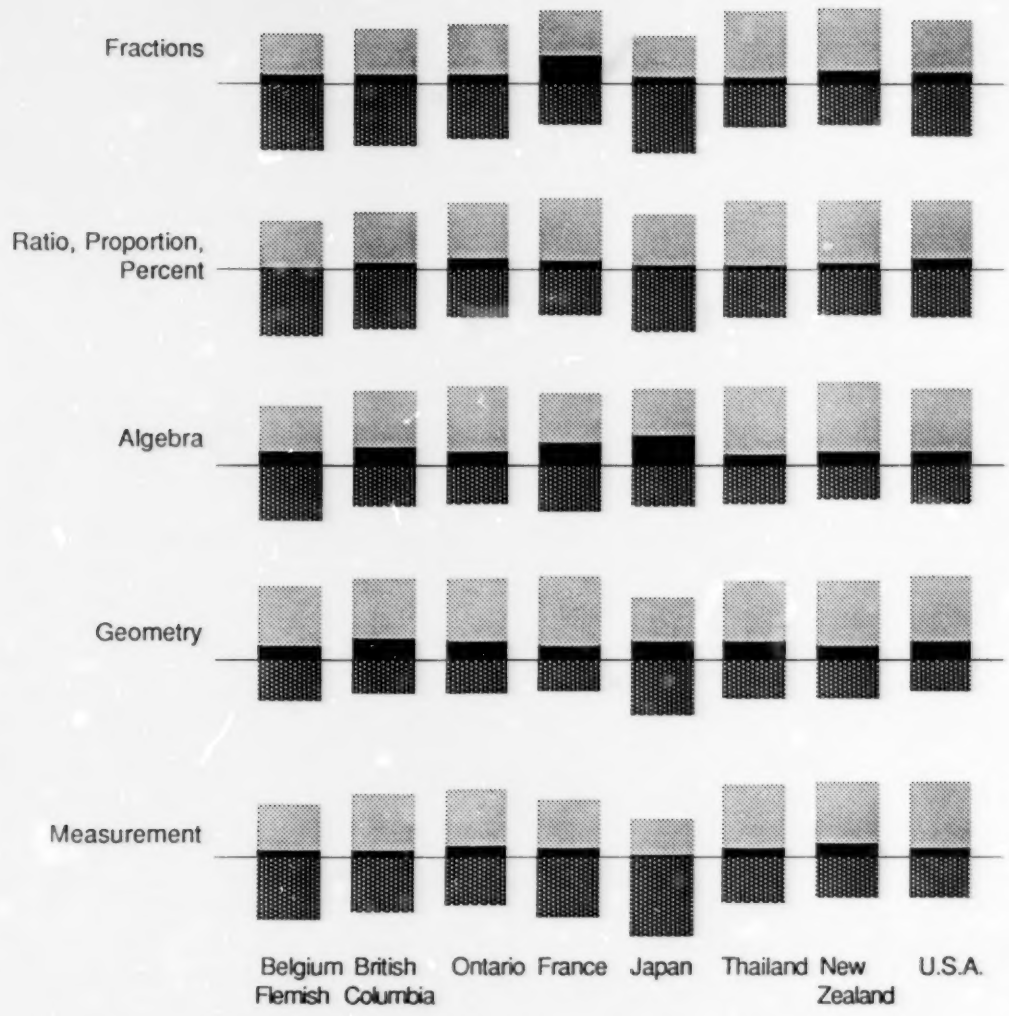
First, it should be pointed out that in the Population A data, there are very minor differences in achievement between boys and girls – pretest, posttest, or gain – except in a few spatial ability items. There are some small attitude differences, primarily in attitudes towards gender and mathematics.

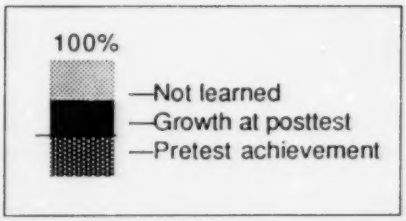
Yet when we look at the SIMS Population B data, taken at the end of secondary school, there are enormous gender differences: for example, only 20-40% of the students are girls. Somewhere between the Population A and B levels, there has been a shift, presumably in both achievement and attitudes. A proper longitudinal study would allow us to follow these changes. Given that in the year after Population A there is, in many countries, the beginning of strong tracking, it would be very important to look at attitude change and cognitive growth over time at about this level.

For a Third IEA Mathematics Study, consideration should be given to a multi-grade design, with pretest and posttest evaluation of achievement and attitudes and careful identification of opportunities and tracking.

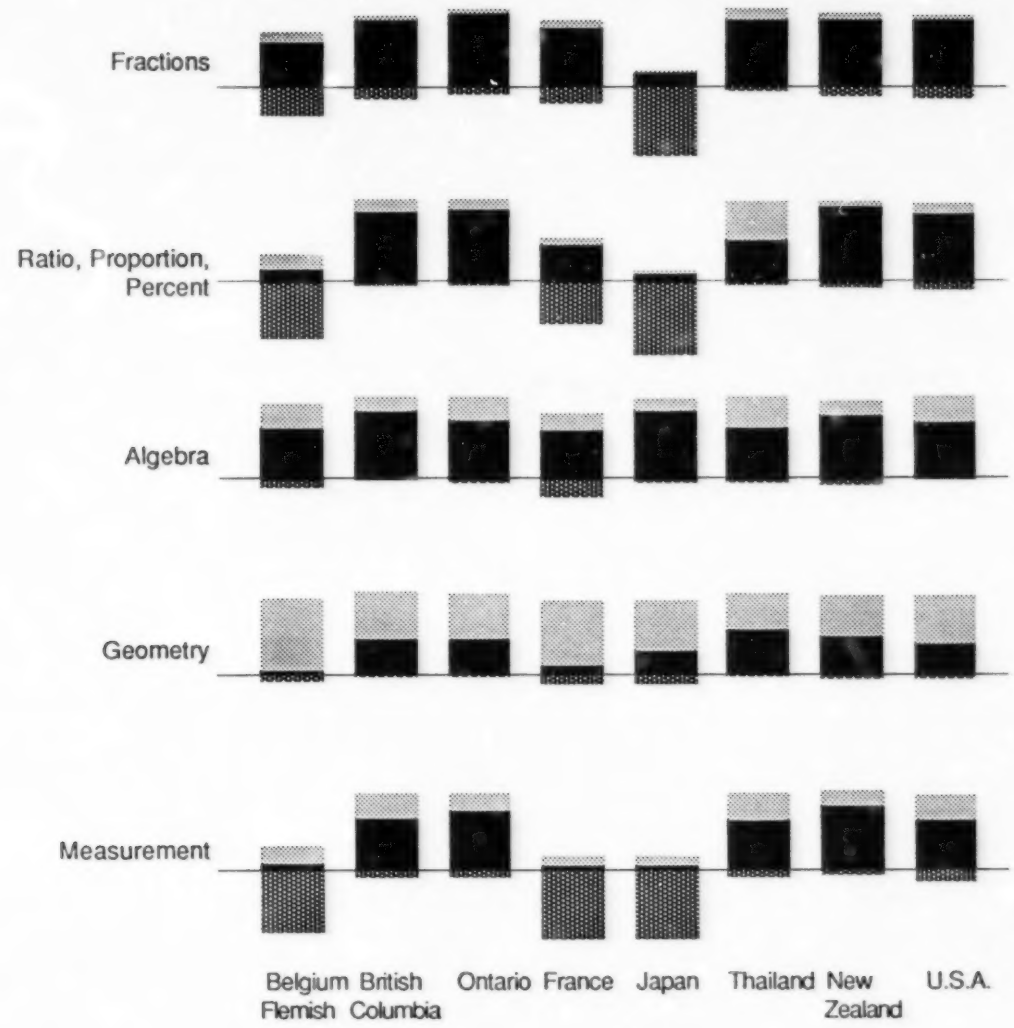
---

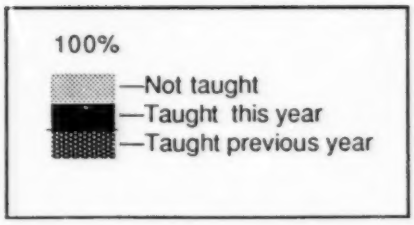
8. In David Walker's original opportunity to learn study, in the IEA Feasibility Project of 1959, teachers were asked whether the mathematics content might be attained outside of school.



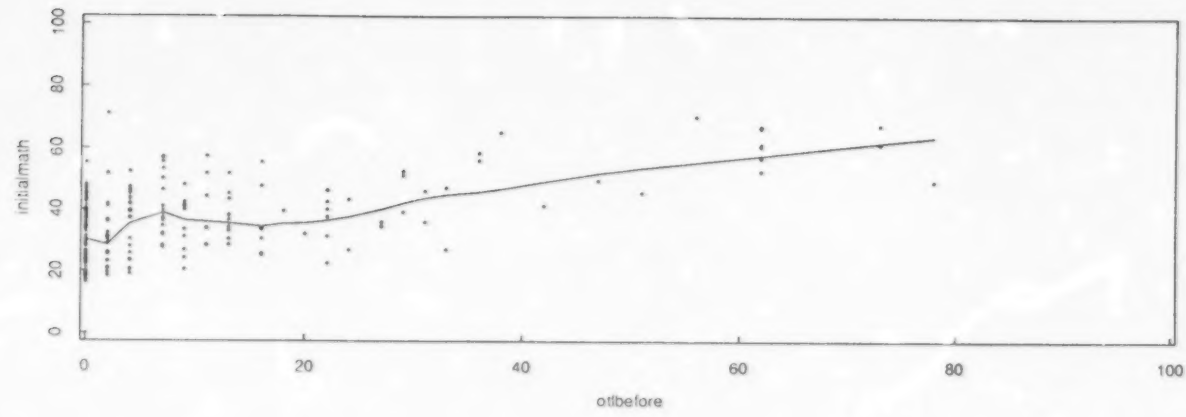


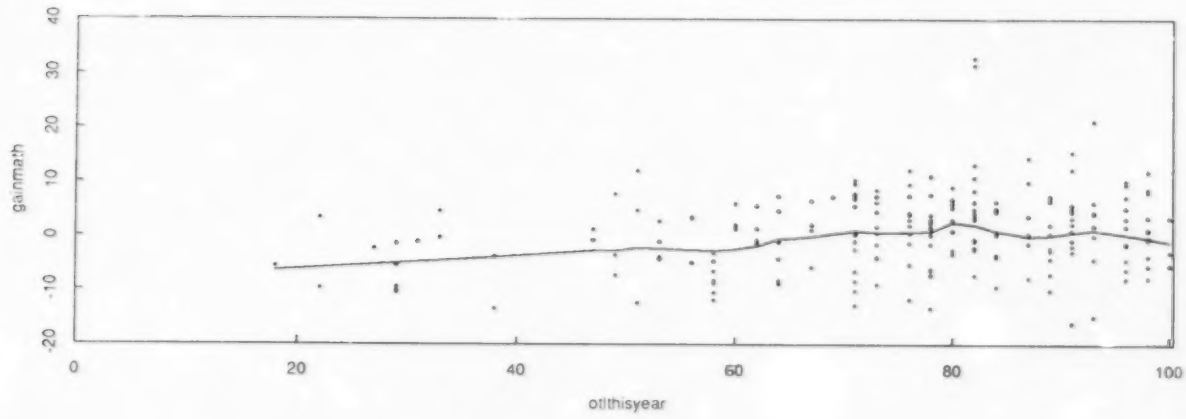
Display 1
Mathematics Achievement and
Growth by Topic and Country

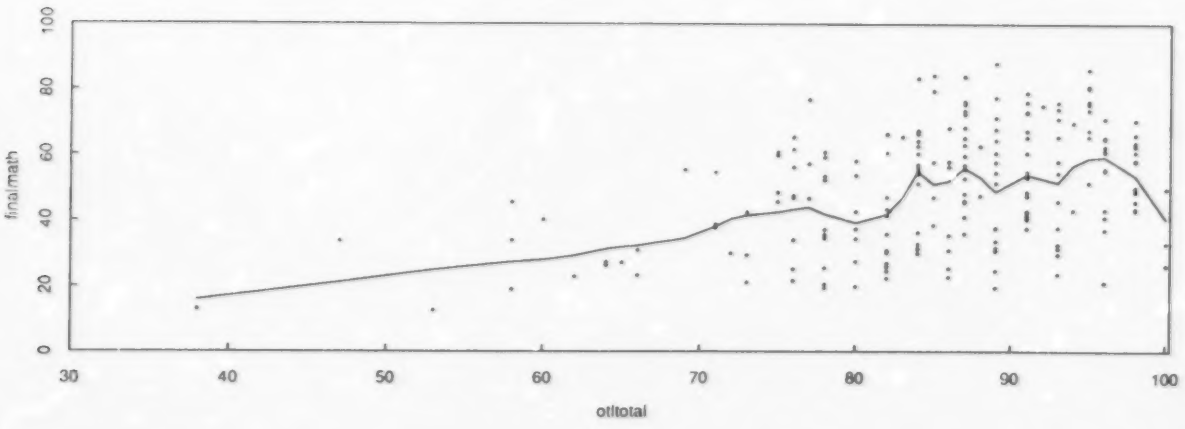




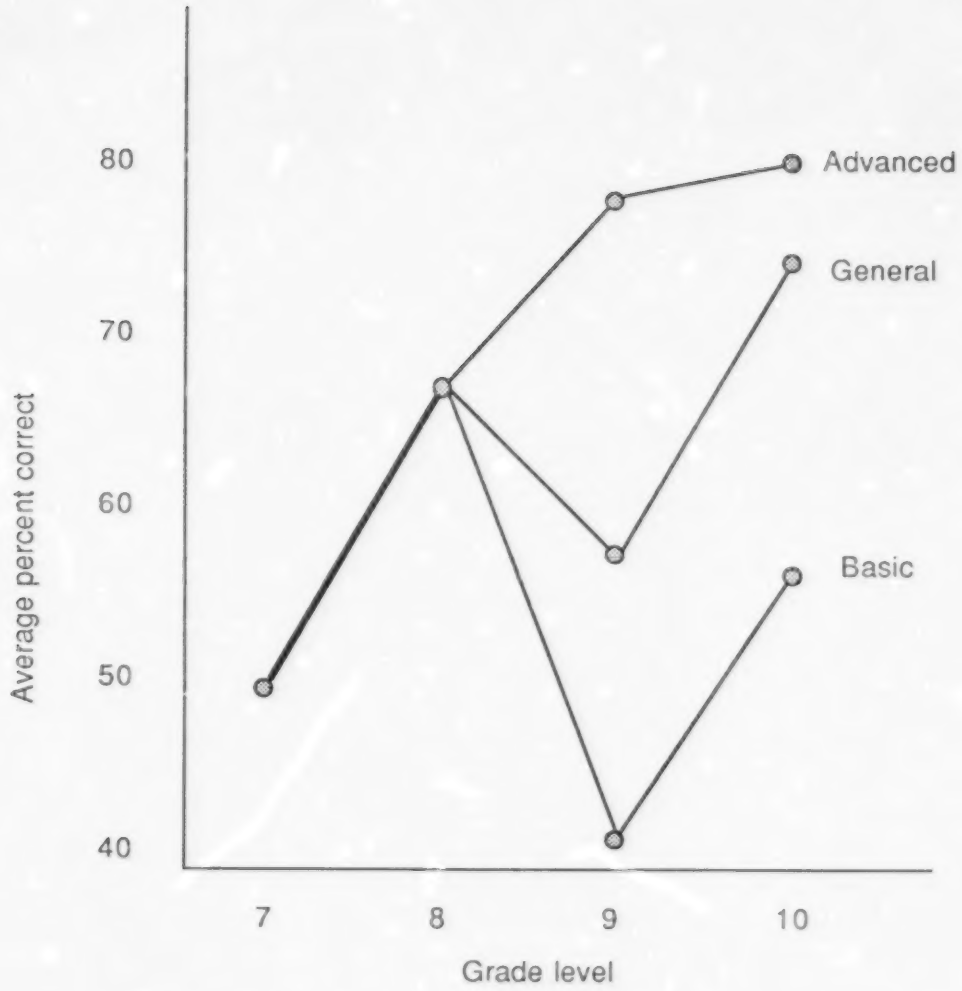
Display 2
Opportunity to Learn by
Topic and Country







Display 3
Pretest, Residualized Gain, and Posttest Achievement by Opportunity to Learn Before, This Year, and Total for Arithmetic in U.S.A. Classrooms



Display 4
Shift in Arithmetic Skill over Grades 7 to 10
in Ontario, with a Division into Tracks
Starting in Grade 9